

Japanese Patent Application Laid-Open No. 07-240722

(43) Publication Date: September 12, 1995

(21) Application Number: Japanese Patent Application No. 6-29336

(22) Filling Date: February 28, 1994

(71) Applicant: 000003078

Toshiba Corp.

(72) Inventor: Shigenobu Minami

[0065]

This first estimates in the delay estimating portion 108 the delay difference $\tau_L - \tau_R$ between the left and right voices on the basis of the output voices $Y_R(z)$ and $Y_L(z)$ of the left and right microphones 101-1 and 101-2 so as to obtain the position information LOCT of the speaker.

[0066]

The microphone array process is simultaneously executed. This aligns in the delay and gain applying portions 102-1 and 102-2 the timings of the output voices $Y_R(z)$ and $Y_L(z)$ of the left and right microphones 101-1 and 101-2 on the basis of the delay difference $\tau_L - \tau_R$ obtained by the delay estimating portion 108, adds the respective outputs $Y_{RD}(z)$ and $Y_{LD}(z)$ of the delay and gain applying portions 102-1 and 102-2 in the adder 103, and enhances the correlation component. In other words, on the assumption that the output of the adder 103 is $Y_{add}(z)$, the relation

$Y_{\text{add}}(z) = Y_{\text{RD}}(z) + Y_{\text{LD}}(z)$ is established.

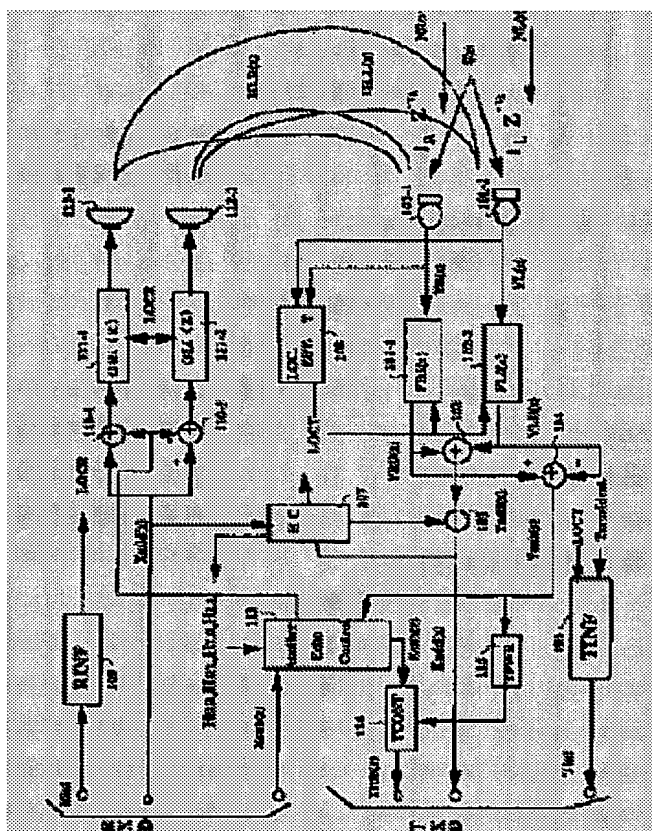
99

†

Abstract of JP7240722

PURPOSE:To obtain the flexible voice system utilizing merits of both of a complete stereo and a pseudo stereo by adding outputs of delay gain giving parts of right and left channels to emphasize the correlation component and obtaining the difference between outputs of right and left giving parts to output it as additional information.

CONSTITUTION:A delay difference $\tau_{\text{auL}}\text{-}\tau_{\text{auR}}$ between right and left voices is estimated from output voices $Y_R(Z)$ and $Y_L(Z)$ of left and right microphones 101-1, 101-2 to obtain position information LOCT of a speaker in a delay estimation part 108. Simultaneously, microphone array processing is performed where outputs $Y_{RD}(Z)$ and $Y_{LD}(Z)$ of delay gain giving parts 102-1 and 102-2 are added by an adder 103 to emphasize the correlation component and the output of an output echo canceller 107 of the adder 103 is subtracted by a subtractor 105 to eliminate the echo component. Second additional information $T_{\text{sub}}(Z)$ is generated and sent besides execution of this pseudo stereo voice encoding processing. That is, the difference between outputs $Y_{RD}(Z)$ and $Y_{LD}(Z)$ of giving parts 102-1 and 102-2 is obtained by a subtractor 104, and an output $Y_{\text{sub}}(Z)$ is sent as second additional information $T_{\text{sub}}(Z)$ through an echo eliminating part 113.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-240722

(43) 公開日 平成7年(1995)9月12日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 B 14/04	Z	9372-5K		
G 1 0 K 15/00				
H 0 4 B 3/20		4229-5K		
H 0 4 N 5/60	Z	9381-5H		
			G 1 0 K 15/ 00	M
			審査請求 未請求	請求項の数 9 O L (全 14 頁)

(21) 出願番号 特願平6-29336

(22) 出願日 平成6年(1994)2月28日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 南 重信

神奈川県川崎市幸区柳町70番地 株式会社

東芝柳町工場内

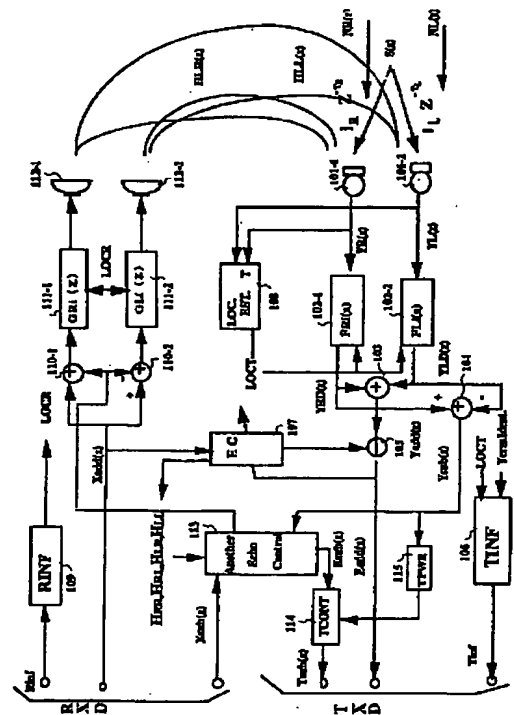
(74) 代理人 弁理士 須山 佐一

(54) 【発明の名称】 音声符号復号化装置、音声符号化装置及び音声復号化装置

(57) 【要約】

【目的】 完全ステレオ・疑似ステレオ両面の利点を生かした柔軟なテレビ電話・マルチメディア端末等の音声系の提供。

【構成】 疑似ステレオ音声情報を構成する左右各チャンネルの和情報 $E_{add}(z)$ と付加情報 T_{inf} に加えて、必要に応じて新たに左右各チャンネル音声の差情報 $T_{sub}(z)$ を符号伝送している。



【特許請求の範囲】

【請求項 1】 相手端末において少なくとも 2 チャンネルの音声より得られた第 1 の複数音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方である第 1 の付加情報と、前記第 1 の複数音声の和または前記第 1 の複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第 1 のモノラル音声と、前記第 1 の複数音声の差または前記第 1 の複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第 2 のモノラル音声とを受信し、前記第 1 のモノラル音声と前記第 1 の付加情報より疑似ステレオ音声を合成する手段と、前記疑似ステレオ音声と前記第 2 のモノラル音声より完全ステレオ音声を合成する手段とを有する復号化部と、

少なくとも 2 チャンネルのマイクより集音した自端末の第 2 の複数音声を用い、第 2 の複数音声間の伝達関数の近似情報と発言者位置情報のうち少なくとも一方である第 2 の付加情報と、前記複数音声の和または前記複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第 3 のモノラル音声と、前記複数音声の差または前記複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第 4 のモノラル音声を生成する符号化部と、

前記第 1 のモノラル音声と前記第 3 のモノラル音声より前記第 3 のモノラル音声に含まれる前記第 1 モノラル音声のエコー成分を抑圧する第 1 のエコー抑圧手段と、前記第 2 のモノラル音声と前記第 4 のモノラル音声より前記第 4 のモノラル音声に含まれる前記第 2 のモノラル音声のエコー成分を抑圧する第 2 のエコー抑圧手段とを具備することを特徴とする音声符号復号化装置。

【請求項 2】 第 1 のエコー抑圧手段は、付加情報を用い過去の複数の推定伝達関数より新たな付加情報に対応する推定伝達関数の初期値を設定する疑似ステレオ用エコーキャンセラ方式が用いられていることを特徴とする請求項 1 記載の音声符号復号化装置。

【請求項 3】 第 2 のエコー抑圧手段は、付加情報を用い過去の複数の推定伝達関数より新たな付加情報に対応する推定伝達関数の初期値を設定する疑似ステレオ用エコーキャンセラ方式が用いられていることを特徴とする請求項 1 記載の音声符号復号化装置。

【請求項 4】 少なくとも 2 チャンネルの音声より前記音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方を得る手段と、前記各音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方の情報を用いて左右の音声に利得または遅延のうち少なくとも一方を与えるマイクアレイ手段と、前記マイクアレイ手段により出力された音声を加算する手段と、前記マイクアレイ手段により出力された音声を減算する手段とを具備することを特徴とする音声符号化装置。

【請求項 5】 少なくとも 2 チャンネルの音声より得られた複数音声間の伝達関数の近似情報と発言者位置情報

の少なくともいずれか一方である付加情報と、前記複数音声の和または前記複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第 1 のモノラル音声より疑似ステレオ音声を合成する手段と、前記複数音声の差または前記複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第 2 のモノラル音声を用いて完全ステレオ音声を合成する手段とを具備することを特徴とする音声復号化装置。

【請求項 6】 少なくとも 2 チャンネルのステレオ音声を入力し、各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を送出することを特徴とするステレオ音声符号化方式。

【請求項 7】 少なくとも 2 チャンネルのステレオ音声を入力し、各ステレオ音声の相関情報に基づいて各ステレオ音声に遅延差・利得差の少なくとも一方の処理を施した後、処理後の各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を送出することを特徴とする請求項 6 記載のステレオ音声符号化方式。

【請求項 8】 受信した各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報をもとにステレオ音声を合成することを特徴とするステレオ音声復号方式。

【請求項 9】 各ステレオ音声の相関情報に基づいて各ステレオ音声に遅延差・利得差の少なくとも一方の処理を施したステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を受信し、これらの情報をもとにステレオ音声を合成することを特徴とするステレオ音声復号方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、例えばテレビ会議システムやマルチメディア端末の音声系に適用される音声符号復号化装置、音声符号化装置及び音声復号化装置に関する。

【0002】

【従来の技術】 図 14 に示すテレビ会議システムや図 15 に示すマルチメディア端末の音声系においては、通話品質の向上やスムーズな通話を実現することが要請されている。 そのためには、発言者の顔位置に音像を定位するステレオ音声技術や複数マイクロフォン出力を合成することにより背景雑音を抑圧するマイクロフォンアレイ技術、スピーカ出力がマイクロフォンに入力することにより発生するエコー感やハウリングを抑圧するエコーキャンセラ (EC) の採用等が有効である。

【0003】 ところで、ステレオ音声を実現するシステムとしては、従来の完全ステレオ音声系の他にいわゆる疑似ステレオ音声系のものがあるが、まず完全ステレオ音声系のシステムに上述した EC の技術を採用した例を図 16 に示す。

3

【0004】同図において、1R は右チャンネルスピーカ、1L は左チャンネルスピーカ、2R は右チャンネルマイク、2L は左チャンネルマイクをそれぞれ示している。右チャンネルスピーカ 1R から出力される受信信号は、エコーキャンセラ E_{CRR} 及びエコーキャンセラ E_{CRL} に入力され、左チャンネルスピーカ 1L から出力される受信信号は、エコーキャンセラ E_{CLR} 及びエコーキャンセラ E_{CLL} に入力される。加算器 3a において、エコーキャンセラ E_{CRR} の出力とエコーキャンセラ E_{CLR} の出力とが合成され、減算器 4a において、右チャンネルマイク 2R より入力された送信信号からこの合成信号が差し引かれる。また、加算器 3b において、エコーキャンセラ E_{CLL} の出力とエコーキャンセラ E_{CRL} の出力とが合成され、減算器 4b において、左チャンネルマイク 2L より入力された送信信号からこの合成信号が差し引かれる。

【0005】このようなシステムでは、通常のステレオ音声のサービスをエコー感やハウリングの抑圧を実現しつつ完全に提供できるが、次のような問題点がある。

【0006】(1) マイクロフオンの数 (=スピーカの数) だけの伝送路を必要とするため、ステレオのチャンネル数が増大すると、通信コストが増加する。

【0007】(2) 複数のスピーカ・マイクロフォンを用いているため、反響路の数が複数となり、これによりエコーキャンセラがマイクロフオンの数とスピーカの数とを乗算した数だけ必要になり、処理量が増加する。

【0008】(3) エコーキャンセラが上述した数だけ必要になるため、決定すべき未知数が増加し、また入力信号間の相関も増大する。このため、エコーキャンセラの収束速度が低下する。

【0009】一方、疑似ステレオ音声系は例えば図 17 に示すように構成される。

【0010】同図に示す疑似ステレオ音声系の受信側では、受信したモノラル音声信号 X_R を、左右で特性の独立な伝達関数 (音像定位関数と呼ぶ。) G_R (Z) 及び G_L (Z) を介して右チャンネルスピーカ 1R 及び左チャ *

$$F_{Ri}(z) = I_{Ri} z^{-Ri}$$

$$F_{Li}(z) = I_{Li} z^{-Li}$$

ここで、疑似ステレオ音声系に適用する音響用エコーキャンセラは、従来のモノラル音声用エコーキャンセラを用いることができる。

【0015】この場合、エコーキャンセラ EC は、図 17 に示すように、受信音声 X_R の伝送路と送信音声 X_S

$$H_i(z) = F_{Ri}(z) (G_{Ri}(z) H_{RR}(z) + G_{Li}(z) H_{LR}(z)) + F_{Li}(z) (G_{Ri}(z) H_{RL}(z) + G_{Li}(z) H_{LL}(z))$$

..... (3)

4

*ネルスピーカ 1L より出力する。音像定位関数 G

_R (Z) 及び G_L (Z) は、モノラル音声信号 X_R と同時に受信した付加情報 d_R 及び予め指定された情報もしくはマウス等によりユーザによって指定された情報に基づいて、聴者に適切なステレオ感を与える (これを疑似ステレオ音像合成処理と呼ぶ。)

【0011】また、送信側では、発言者位置を推定することにより、音源位置情報を得て付加情報 d_S として送信モノラル音声 X_S とともに送信する (これを音源位置推定処理と呼ぶ。)。なお、この音源位置情報を用いて発言者の口元に指向性を向けるマイクアレイ処理も容易に併用することができる。

【0012】上述した疑似ステレオ音像合成処理は、(1) 式に示す音像定位関数 G_R (Z) 及び G_L (Z) により左右の耳に到達する直接波音声の利得差・遅延差を合成して聴者に特定の音像定位を与えるものである。

【0013】

【数 1】

$$G_R(z) = g_R z^{-rR}$$

$$G_L(z) = g_L z^{-rL}$$

..... (1)

また、マイクアレイ処理は、左右の各マイク 2R、2L の出力を特性の独立な伝達関数 (マイク出力合成関数と呼ぶ。) F_{Ri} (Z) 及び F_{Li} (Z) を介した後合成することにより、特定の発言者から到達する音声のみを強調して雑音を抑圧する。ここで、マイク出力合成関数 F_{Ri} (Z) 及び F_{Li} (Z) は、通常、(2) 式を用いて左右のマイクの出力に適切な利得、遅延を与えることにより特定の音源位置から到達する音声を強調するように設定する。

【0014】

【数 2】

..... (2)

の伝送路との間に挿入される。

【0016】また、エコーキャンセラから見た総合的な反響路特性 H(z) は、

【数 3】

となる。

【0017】(3)式から明らかなように、スピーカ・マイク間伝達関数 $H_{RR}(z)$ 、 $H_{LL}(z)$ 、 $H_{RL}(z)$ 、 $H_{LR}(z)$ は、音響特性が定常な限り一定であるが、音像定位関数や音声合成関数は、音像や発言者位置が変化すると、その都度変動する。ここで、図17に示したように単一のエコーキャンセラECを適用すると、反響路特性が頻繁に変動することになり、打ち消し量が劣化する。

【0018】本発明者は、この問題点を改善する手法として、いわゆる分解合成形エコーキャンセラを提唱した。この分解合成形エコーキャンセラの概略を図18に基づいて説明する。

【0019】まず、この分解合成形エコーキャンセラでは、音像定位関数・マイク出力合成関数の両方が定常な*

$$\hat{H}_i(z) = L_i(z) \bar{H}(z)$$

ここで

$$\hat{H}_i(z) = (\hat{H}_i(z), \hat{H}_{i-1}(z), \hat{H}_{i-2}(z), \hat{H}_{i-3}(z))^T$$

$$\bar{H}(z) = (\hat{H}_{RR}(z), \hat{H}_{RL}(z), \hat{H}_{LR}(z), \hat{H}_{LL}(z))^T$$

$$L_i(z) =$$

$$\begin{bmatrix} F_{RI}(z) G_{RI}(z) & F_{LI}(z) G_{RI}(z) & F_{RI}(z) G_{LI}(z) & F_{LI}(z) G_{LI}(z) \\ F_{RI-1}(z) G_{RI-1}(z) & F_{LI-1}(z) G_{RI-1}(z) & F_{RI-1}(z) G_{LI-1}(z) & F_{LI-1}(z) G_{LI-1}(z) \\ F_{RI-2}(z) G_{RI-2}(z) & F_{LI-2}(z) G_{RI-2}(z) & F_{RI-2}(z) G_{LI-2}(z) & F_{LI-2}(z) G_{LI-2}(z) \\ F_{RI-3}(z) G_{RI-3}(z) & F_{LI-3}(z) G_{RI-3}(z) & F_{RI-3}(z) G_{LI-3}(z) & F_{LI-3}(z) G_{LI-3}(z) \end{bmatrix}$$

本方式は、このような関係を利用して、単一のエコーキャンセラECで得た複数過去の推定伝達関数をもとに(5)式に示すスピーカ・マイク間伝達関数を得る。 ※

$$\bar{H}(z) = L_i^{-1}(z) \hat{H}_i(z) \quad \text{..... (5)}$$

そして、 $i+1$ 区間で新たな音像定位関数・マイク出力合成関数が得られたら、

$$\bar{H}_{i+1}(z) = L_{i+1}(z) \hat{H}(z) \quad \text{..... (6)}$$

$$\text{但し、} L_{i+1}(z) = [F_{RI+1}(z) G_{RI+1} \quad F_{RI+1}(z) G_{RI+1} \quad F_{LI+1}(z) G_{RI+1} \quad F_{LI+1}(z) G_{LI+1}]^T$$

なる演算により、 $i+1$ 区間のエコーキャンセラECの係数の初期値を得ることができる。

【0023】

【発明が解決しようとする課題】 このように疑似ステレ 50

*区間を単位とした i th区間を考える。

【0020】ここで、過去4区間のエコーキャンセラEC(図18)の推定伝達関数を

【数4】

$$\hat{H}_{i-3}(z), \hat{H}_{i-2}(z), \hat{H}_{i-1}(z), \hat{H}_i(z)$$

とすると、反響路が定常で、推定が正確であると仮定すれば、この推定特性は、音像定位関数・マイク出力合成関数とは独立なスピーカ・マイク間伝達関数

【数5】

$$\hat{H}_{RR}(z), \hat{H}_{RL}(z), \hat{H}_{LR}(z), \hat{H}_{LL}(z)$$

を用いて以下のように表わすことができる。

【0021】

【数6】

$$\text{..... (4)}$$

※【0022】

【数7】

オ音声系には多くの利点があるが、実際のテレビ会議システムやマルチメディア端末に应用するためには以下のような課題が依然として残っている。

【0024】(1) 疑似ステレオ方式は単独発言を前提

としているため、ポイント・ポイント接続のテレビ会議システムやマルチメディア端末では複数の話者が同時に発言したときに、またマルチポイント接続のテレビ会議システムやマルチメディア端末では複数端末で発言が同時に行われたときに、音像の乱れが発生する。

【0025】また、本発明者が提唱した分解合成形エコーキャンセラも、疑似ステレオを前提としているため、複数発言時の対応がなされていない。

【0026】(2) 通常の使用条件では相手端末が疑似ステレオ方式を採用しているとは限らないため、この場合は相互接続が困難になる。

【0027】すなわち、疑似ステレオ、完全ステレオの両方に柔軟に対応できる音声系の採用が望ましい。

【0028】(3) 上述した分解合成形エコーキャンセラにおいては、(5) 式及び(6) 式の係数分解・合成処理は利得制御のみの場合は容易であるが、遅延制御の場合には一種のフィルタ処理になるため、系の安定性や実現性が困難になる場合がある。(4) 音像定位関数の変化はマイク合成関数の変化とは異なり、反響路特性の遅延成分により、過渡的な状態を生成する。このとき、エコーキャンセラの打ち消し量が一時的に劣化する。

【0029】そこで、本発明は、上述の課題を解決するため、完全ステレオ・疑似ステレオ両面の利点を生かした柔軟なテレビ電話・マルチメディア端末等の音声系を提供することを目的としている。

【0030】より具体的には、本発明の第1の目的は、音源が複数発言の場合にも効率的なステレオ音声符号化を実現できる音声符号復号化装置、音声符号化装置及び音声復号化装置を提供することにある。

【0031】本発明の第2の目的は、疑似ステレオ、完全ステレオの両方に柔軟に対応できる音声符号復号化装置、音声符号化装置及び音声復号化装置を提供することにある。

【0032】本発明の第3の目的は、分解合成形エコーキャンセラにおいて系の安定性や実現性を容易にすることができる音声符号復号化装置を提供することにある。

【0033】本発明の第4の目的は、エコーキャンセラの打ち消し量の一時的な劣化を防止することができる音声符号復号化装置を提供することにある。

【0034】

【課題を解決するための手段】かかる課題を解決するため、請求項1記載の本発明は、相手端末において少なくとも2チャンネルの音声より得られた第1の複数音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方である第1の付加情報と、前記第1の複数音声の和または前記第1の複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第1のモノラル音声と、前記第1の複数音声の差または前記第1の複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第2のモノラル音声とを受信し、前記第1のモノラル

音声と前記第1の付加情報より疑似ステレオ音声を作成する手段と、前記疑似ステレオ音声と前記第2のモノラル音声より完全ステレオ音声を作成する手段とを有する復号化部と、少なくとも2チャンネルのマイクより集音した自端末の第2の複数音声を用い、第2の複数音声間の伝達関数の近似情報と発言者位置情報のうち少なくとも一方である第2の付加情報と、前記複数音声の和または前記複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第3のモノラル音声と、前記複数音声の差または前記複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第4のモノラル音声を生成する符号化部と、前記第1のモノラル音声と前記第3のモノラル音声より前記第3のモノラル音声に含まれる前記第1モノラル音声のエコー成分を抑圧する第1のエコー抑圧手段と、前記第2のモノラル音声と前記第4のモノラル音声より前記第4のモノラル音声に含まれる前記第2のモノラル音声のエコー成分を抑圧する第2のエコー抑圧手段とを具備する。

【0035】請求項2記載の本発明は、請求項1記載の音声符号復号化装置において、前記第1のエコー抑圧手段は、付加情報を用い過去の複数の推定伝達関数より新たな付加情報に対応する推定伝達関数の初期値を設定する疑似ステレオ用エコーキャンセラ方式が用いられている。

【0036】請求項3記載の本発明は、請求項1記載の音声符号復号化装置において、前記第2のエコー抑圧手段は、付加情報を用い過去の複数の推定伝達関数より新たな付加情報に対応する推定伝達関数の初期値を設定する疑似ステレオ用エコーキャンセラ方式が用いられている。

【0037】請求項4記載の本発明は、少なくとも2チャンネルの音声より前記音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方を得る手段と、前記各音声間の伝達関数の近似情報または発言者位置情報のうち少なくとも一方の情報を用いて左右の音声に利得または遅延のうち少なくとも一方を与えるマイクアレイ手段と、前記マイクアレイ手段により出力された音声を加算する手段と、前記マイクアレイ手段により出力された音声が減算する手段とを具備する。

【0038】請求項5記載の本発明は、少なくとも2チャンネルの音声より得られた複数音声間の伝達関数の近似情報と発言者位置情報のうち少なくともいずれか一方である付加情報と、前記複数音声の和または前記複数音声に遅延もしくは利得を与えた後の和成分のいずれかである第1のモノラル音声より疑似ステレオ音声を作成する手段と、前記複数音声の差または前記複数音声に遅延もしくは利得を与えた後の差成分のいずれかである第2のモノラル音声を用いて完全なステレオ音声を作成する手段とを具備する。

【0039】請求項6記載の本発明は、少なくとも2チ

チャンネルのステレオ音声を入力し、各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を送出することを特徴とする。

【0040】請求項7記載の本発明は、請求項6記載のステレオ音声符号化方式において、少なくとも2チャンネルのステレオ音声を入力し、各ステレオ音声の相関情報に基づいて各ステレオ音声に遅延差・利得差の少なくとも一方の処理を施した後、処理後の各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を送出することを特徴とする。

【0041】請求項8記載の本発明は、受信した各ステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報をもとにステレオ音声を合成することを特徴とする。

【0042】請求項9記載の本発明は、各ステレオ音声の相関情報に基づいて各ステレオ音声に遅延差・利得差の少なくとも一方の処理を施したステレオ音声の和成分、差成分、およびステレオ音声間の相関情報もしくは、その近似情報を受信し、これらの情報をもとにステレオ音声を合成することを特徴とする。

【0043】

【作用】本発明では、疑似ステレオ音声情報を構成する左右各チャンネルの和情報と付加情報に加えて、必要に応じて新たに左右各チャンネル音声の差情報を符号伝送することで、音源が複数発言の場合にも効率的なステレオ音声符号化を実現し、また疑似ステレオ、完全ステレオの両方に柔軟に対応できる音声系を実現している。

【0044】また、本発明では、推定係数分解処理において、過去の複数の推定伝達関数からスピーカ・マイク間伝達関数を推定する処理が必要となるが、このとき、推定係数分解用のフィルタに入力する推定伝達関数のインパルス応答を遅延量の大きい応答から少ない応答の順に処理することにより、推定係数分解用のフィルタの安定条件を満たし、これにより分解合成エコーキャンセラにおいて系の安定性や実現性を容易にすることを可能にしている。

【0045】さらに、本発明では、推定伝達関数分解処理により得た新たな伝達関数のインパルス応答を疑似ステレオの音像制御情報とマイクアレイ情報の変化時点で同時にエコーキャンセラの初期値としてセットするのでなく、遅延量の少ない係数より順次、置き換えることにより、エコーキャンセラの打ち消し量の一時的な劣化を防止している。

【0046】

【実施例】以下、発明の実施例の詳細を図面に基づき説明する。

【0047】図1は本発明が適用されるマルチメディア会議システムの構成を示す図である。同図に示すシステムでは、複数のマルチメディア端末TE、…がローカルエリアネットワークLANを介して接続されている。

【0048】マルチメディア端末TEは、大別すると音声系と映像系と静止画像系と伝送系とから構成される。

【0049】これら各系は、以下のように実現される。

【0050】(1) 音声系

音声系は、疑似ステレオを基本としたマルチポイント音声会議を可能にするものである。

【0051】また、複数のウィンドウに応じた音像制御、すなわち各端末に応じたウィンドウ及びウィンドウ内の発言者口元に音像を定位することを可能にする。

10 【0052】マルチポイント音声制御は、各端末から伝送された音声パケットを自端末で合成する分散制御形とする。

【0053】(2) 映像系

LANから送られてきた映像パケットを対応するウィンドウに表示する。

【0054】パケットには映像符号化方式の種類が付与されているため、これに対応する復号方式を適用する。

【0055】また、対応する音声パケットへのポインタを付与することにより、音声と映像の同期も確保する。

20 【0056】さらに、疑似ステレオ情報が発言者の位置情報も含んでいるため、符号化側では、発言者にズームを向けるカメラ制御や発言者のみ高品質に符号化する映像符号化制御が可能になり、復号側では発言者のみカラー表示したり動画像符号化することにより符号化処理の簡略化が可能になる。

【0057】(3) 静止画像系

図2に示すように書類は画像スキャナで読み取るとともに、必要に応じて計算機のファイルとして蓄積する。

30 【0058】また、処理した書類は書類処理部で廃棄するか書類としてファイリングするか決定し、各々の書類蓄積BOXに出力する。

【0059】計算機上の書類は、プリンタにより紙書類として出力される。

【0060】(4) 伝送系

伝送はパケットまたはセル形式を基本とする。

【0061】ここで音声および映像のパケット(セル)は映像、音声情報及び宛先情報以外に自端末の識別番号、発言者の位置情報及び画像、音声の時間関係を規程するタイミング情報を含む。

40 【0062】さて、このように構成されたマルチメディア会議システムにおいて、本発明は以下のように適用される。

【0063】まず、ステレオー疑似ステレオーモノラル音声間の通信互換性を確保する一実施例について説明する。

【0064】図3はその実施例に係る音声系の構成を示す図であり、左右の音声の和情報である主情報と遅延差・利得差ある付加情報からなる疑似ステレオ音声情報に加えて、左右の音声の差情報である第2の付加情報を用いることにより、かかる通信互換性を確保している。

【0065】これは、まず遅延推定部108において、左右のマイク101-1、101-2の出力音声 $Y_R(z)$ 、 $Y_L(z)$ から左右音声間の遅延差 $\tau_L - \tau_R$ を推定して発言者の位置情報LOCTを得る。

【0066】同時に、マイクアレイ処理を行う。これは、遅延・利得付与部102-1、102-2において、遅延推定部108で得られた遅延差 $\tau_L - \tau_R$ に基づき左右のマイク101-1、101-2の出力音声 $Y_R(z)$ 、 $Y_L(z)$ のタイミングを合わせ、遅延・利得付与部102-1、102-2の各出力 $Y_{RD}(z)$ 、 $Y_{LD}(z)$ を加算器103において加算し、相関成分を強調する。つまり、加算器103の出力を $Y_{add}(z)$ とすると、

$$Y_{add}(z) = Y_{RD}(z) + Y_{LD}(z)$$

となる。

【0067】そして、減算器105において、加算器103の出力(モノラル音声) $Y_{add}(z)$ から音響用エコーキャンセラ107の出力が減算され、エコー成分が除去される。この減算器105の出力 $E_{add}(z)$ は、端末識別子TermIdentと発言者位置情報LOCTから付加情報処理部106で生成される端末識別情報 T_{inf} とともに疑似ステレオ音声の送信情報TXDとなる。

【0068】本発明では、このような疑似ステレオ音声符号化処理に加え、第2の付加情報 $T_{sub}(z)$ を生成して送出している。この第2の付加情報 $T_{sub}(z)$ は次のように生成される。

【0069】まず、減算器104において、遅延・利得付与部102-1、102-2の各出力 $Y_{RD}(z)$ 、 $Y_{LD}(z)$ の差分が求められる。つまり、減算器104の出力 $Y_{sub}(z)$ とすると、

$$Y_{sub}(z) = Y_{RD}(z) - Y_{LD}(z)$$

となる。

【0070】この減算器104の出力 $Y_{sub}(z)$ は、第2のエコー除去部113を経由し完全ステレオ音声用の第2の付加情報 $T_{sub}(z)$ として送出される。

【0071】但し、 $Y_{RD}(z)$ 、 $Y_{LD}(z)$ の差成分の電力は、電力検出器115により検出されており、その差成分の電力が小さい場合、すなわち、左右のマイク出力音声間の相関が強い場合には、送信制御部114により、この第2の付加情報の送出が禁止されるようになっている。

【0072】なお、上述した遅延・利得付与を省略することもでき、これにより構成の簡略化が可能である。

【0073】一方、受信側では、送られて来た疑似ステレオ情報であるモノラル音声 $Y_{add}(z)$ と付加情報 R_{inf} とにより音像を与える疑似ステレオ音声を合成する。

【0074】また、第2の付加情報である差成分を用いることにより完全ステレオ音声を合成できる。

【0075】すなわち、加算器110-1において、 $X_R(z) = Y_{add}(z) + Y_{sub}(z)$

を求め、減算器110-2において、 $X_L(z) = Y_{add}(z) - Y_{sub}(z)$ を求める。

【0076】ここで、 $Y_{sub}(z)$ の電力が小さいとき、または $Y_{sub}(z)$ を受信しなかったときには、 $Y_R(z) = Y_L(z)$ となり完全ステレオ音声となる。

【0077】疑似ステレオ合成部111-1、111-2では、受信付加情報 R_{inf} より得た発言者位置情報LOCRを用いてモノラル音声に利得・遅延を与えて疑似ステレオ音声を合成する。また、完全ステレオ入力時にはマイクアレイ処理で調整した利得・遅延を元に戻す。

【0078】従って、本発明によれば、受信音声情報がモノラル音声の場合には疑似ステレオ化しないことによりモノラル音声として再生できる。また、付加情報が得られたときには通常の疑似ステレオ音声合成でき、さらに第2の付加情報が得られたときには、完全ステレオ音声として合成できる。

【0079】ところで、図3に示した推定関数分解・合成形エコーキャンセラ107は、すでに明らかなようにモノラル音声を用いる疑似ステレオ方式を前提としている。このため、完全ステレオ性が強い場合、すなわち $X_{sub}(z)$ の信号電力が大きい場合にはこの信号成分のエコーを除去することができない。

【0080】そこで、本発明では、完全ステレオ用の付加情報、すなわちマイク出力の差成分である $X_{sub}(z)$ 、 $Y_{sub}(z)$ に第2のエコー除去部113を適用し、この成分のエコーを除去する。

【0081】第2のエコー除去部113としては、図4に示すような音声スイッチ方式を用いても良いし、図5に示すような推定関数分解・合成形エコーキャンセラを用いても良い。

【0082】次に、分解合成形エコーキャンセラにおいて系の安定性や実現性を容易にする一実施例について説明する。

【0083】従来の技術で説明した(5)(6)式の係数分解・合成処理は利得制御のみの場合は容易であるが、遅延制御の場合には一種のフィルタ処理になるため、系の安定性及び実現性に留意する必要がある。ここでは、図6に示すようにマイクアレイに遅延制御を適用する場合を想定して説明する。

【0084】マイク出力制御関数 $F_{Ri}(z)$ 、 $F_{Li}(z)$ は、例えば遅延推定(Delay Est.)による発言者位置検出で得たith. 定常区間の左右のマイク出力信号の利益差 $l_i (= \sigma_R / \sigma_L$ 、但し σ_R 、 σ_L は各々左右マイク出力信号のレベル)遅延(位相)差 $\tau_i (= \tau_R - \tau_L$ 、但し τ_R 、 τ_L は発言者の口元から左右のマイクまでの遅延量(sample))を用い

$$F_{Ri}(z) = l_{Ri} Z^{-\tau_{Ri}}$$

$$F_{Li}(z) = l_{Li} Z^{-\tau_{Li}} \dots (7)$$

なる演算を行う。

【0085】このような演算処理により、特定位置の発
言者音声を効率良く合成できる。

【0086】ここで、 i 、 $i-1$ th. 定常区間の推定伝達*

*関数を用いてスピーカ・マイク間の伝達関数を(5)式
に基づいて求めると以下の結果を得る。

【0087】

【数8】

$$\hat{H}_R(z) = (I_{Li-1} z^{-Li-1} \hat{H}_i(z) - I_{Li} z^{-Li} \hat{H}_{i-1}(z)) / D$$

$$\hat{H}_L(z) = (-I_{Ri-1} z^{-Ri-1} \hat{H}_{i-1}(z) + I_{Ri} z^{-Ri} \hat{H}_i(z)) / D$$

$$\text{但し、} D = I_{Ri} I_{Li-1} z^{-(Ri+Li-1)} - I_{Ri-1} I_{Li} z^{-(Ri-1+Li)}$$

これらの結果より、図7に示すように、スピーカ・マイ
ク間伝達関数は過去2定常区間 i 、 $i-1$ のエコーキャン
セラの推定伝達関数を $F_{Ai}(z)$ 、 $F_{Bi}(z)$ なる特性
を有するフィルタに通すことにより得られることが分 ※

※。

【0088】図7より $F_{Ai}(z)$ 、 $F_{Bi}(z)$ は、

【数9】

$$F_{Ai}(z) = \frac{\alpha_{Bi} z^{-\tau_{Bi}}}{1 - \alpha_{Ai} z^{-\tau_{Ai}}}, \quad F_{Bi}(z) = \frac{\alpha_{Ci} z^{-\tau_{Ci}}}{1 - \alpha_{Ai} z^{-\tau_{Ai}}}$$

ここで、 α_{Ai} 、 α_{Bi} 、 α_{Ci} 、 τ_{Ai} 、 τ_{Bi} 、 τ_{Ci} の各パラメータは以下のように2つの形式で
表わすことができる。

(CASE-1)

$$\alpha_{Ai} = \frac{I_{Ri-1} I_{Li}}{I_{Ri} I_{Li-1}}, \quad \alpha_{Bi} = \frac{I}{I_{Ri}}, \quad \alpha_{Ci} = \frac{-I_{Li}}{I_{Ri} I_{Li-1}}$$

$$\tau_{Ai} = \tau_{Ri-1} - \tau_{Ri} + \tau_{Li} - \tau_{Li-1}, \quad \tau_{Bi} = -\tau_{Ri}, \quad \tau_{Ci} = \tau_{Li} - \tau_{Ri} - \tau_{Li-1}$$

(CASE-2)

$$\alpha_{Ai} = \frac{I_{Ri} I_{Li-1}}{I_{Ri-1} I_{Li}}, \quad \alpha_{Bi} = \frac{-I_{Li-1}}{I_{Ri-1} I_{Li}}, \quad \alpha_{Ci} = \frac{I}{I_{Ri-1}}$$

$$\tau_{Ai} = \tau_{Ri} - \tau_{Ri-1} + \tau_{Li-1} - \tau_{Li}, \quad \tau_{Bi} = -\tau_{Li-1} - \tau_{Ri-1} - \tau_{Li}, \quad \tau_{Ci} = -\tau_{Ri-1}$$

これらのパラメータをマイク入力信号間の利得差 κ_i 、遅延差 τ_i で書きなおすと、

(CASE-1)

$$\alpha_{Ai} = \frac{I_i^2}{I_{i-1}^2}, \quad \alpha_{Bi} = \frac{I}{\kappa I_i}, \quad \alpha_{Ci} = \frac{I_{i-1}}{\kappa I_i^2}$$

$$\tau_{Ai} = \tau_{i-1} - \tau_i, \quad \tau_{Bi} = \tau_i / 2 - \tau_{\max}, \quad \tau_{Ci} = \tau_i - \tau_{\max} - \tau_{i-1} / 2$$

(CASE-1)

$$\alpha_{Ai} = \frac{I_i^2}{I_{i-1}^2}, \quad \alpha_{Bi} = \frac{-I_i}{\kappa I_{i-1}^2}, \quad \alpha_{Ci} = \frac{I}{\kappa I_{i-1}}$$

$$\tau_{Ai} = \tau_{i-1} - \tau_i, \quad \tau_{Bi} = \tau_{i-1} - \tau_{\max} - \tau_i / 2, \quad \tau_{Ci} = \tau_{i-1} / 2 - \tau_{\max}$$

ここで、分解フィルタ $F_{Ai}(z)$ 、 $F_{Bi}(z)$ が安定で
ある条件は、 $\alpha_{Ai} < 1$ であり、CASE-1では利得差
が $i-1$ 定常区間から i 定常区間で減少する必要がある。
また、CASE-2では利得差が増加する必要がある。一
方、マイクアレイではマイク出力レベルに比例して利得
を上げる必要がある。従って、遅延差が大きい場合には

左右のマイク出力レベル比も大きいことより、遅延差と
利得差は比例すると考えられる。このとき、CASE-
1、CASE-2とも τ_{Ai} は負になるため、図7のフィル
タはそのままでは実現不可能になる。これを解決するた
め、フィルタ $F_{Ai}(z)$ 、 $F_{Bi}(z)$ に入力するインパ
ルス応答の時間方向を逆に処理する。すなわち、推定伝

達関数、分解フィルタの特性

【数10】

$$\hat{H}_i(z), \hat{H}_{i-1}(z), F_{Ai}(z), F_{Ai}(z)$$

を複素変換して、

$$\hat{H}_i(z^{-1}), \hat{H}_{i-1}(z^{-1}), F_{Ai}(z^{-1}), F_{Ai}(z^{-1})$$

なる特性にした後、通常のフィルタ処理を施す。この処理は、図8に示すように入力する推定伝達関数の時間順序を逆にすることで、図7と同一構成のフィルタ処理で実現できる。

【0089】次に、過渡時の打ち消し量劣化の防止に関する一実施例について説明する。

【0090】音像定位関数の変化は、マイク合成関数の変化とは異なり、反響路特性の遅延成分により、過渡的な状態を生成する。すなわち、音像定位関数の変化から*

$$\begin{aligned} h_{kj}(z) &= F_{Ri+1}(z) G_{Ri+1}(z) h_{Rk} z^{-1} \\ &+ F_{RLi+1}(z) G_{RLi+1}(z) h_{RLk} z^{-1} + F_{Lri+1}(z) G_{Lri+1}(z) h_{LRk} z^{-1} \\ &+ F_{LLi+1}(z) G_{LLi+1}(z) h_{LLk} z^{-1} \quad \dots (k \geq j) \end{aligned}$$

$$\begin{aligned} h_{kj}(z) &= F_{Ri+1}(z) G_{Ri+1}(z) h_{Rk} z^{-1} \\ &+ F_{RLi+1}(z) G_{RLi+1}(z) h_{RLk} z^{-1} + F_{Lri+1}(z) G_{Lri+1}(z) h_{LRk} z^{-1} \\ &+ F_{LLi+1}(z) G_{LLi+1}(z) h_{LLk} z^{-1} \quad \dots (k \geq j) \end{aligned}$$

..... (9)

以上のもとで、 $i+1$ 定常区間、 j サンプル目のエコーキャンセルの伝達関数を、

$$\hat{H}_{i,j}(z) = \sum_{k=1}^N h_{k,j}(z) z^{-k}$$

で与えることにより、音像定位関数の変化時点の一時的な打ち消し量の劣化を防ぐことができる。

【0092】ところで、広帯域ステレオ音声は将来のマルチメディア通信において臨場感・自然感の向上に有効である。この場合、通信路のB-ISDN網は、セルと呼ばれる固定長パケットを用いるため、広帯域ステレオ音声を効率良くパケット符号化する必要が有る。一方、ステレオ音声は上述したように左右の音声間の相関成分を利用して効率良く符号化できる。このような背景から、以下に広帯域ステレオ音声パケット符号化方式およびシステムを提案する。

【0093】ステレオ音声は左右の音声間の相関成分を利用して効率よく符号化できる (S. MINAMI, "A Stereophonic Voice Coding Method For Teleconferencing."

* j サンプル後 ($j < N$ N はエコーキャンセラのタップ数) では、例えば右スピーカ・右マイク間の反響路特性を N タップの FIR フィルタでモデル化して考えると、その前半 j タップ分のデータは変化後の音像定位関数 $G_{Ri}(z)$ の出力であり、後半 $N-j+1$ タップ分は変化以前の $G_{Ri}(z)$ の出力データが残っていることになる。従って、マイク合成関数の変化と同時にエコーキャンセラの全係数を置き換えると、反響路特性インパルス応答の後半は変化以前の特性であるため、一時的な打ち消し量の劣化をきたす。これを防止するには、音像定位関数の変化時に以下に示すようなエコーキャンセラのタップ係数の前半・後半で異なる係数置き換え処理を適用する。例えば (9) 式に示すように、音像定位関数の変化から j サンプルにおける k 番目のエコーキャンセラのタップ係数を考える。

【0091】

【数11】

※【数12】

..... (10)

IEEE ICC '86 46.6, June 1986 及び S. MINAMI, "Stereophonic ADPCM Speech Coding Method" IEEE ICASSP 90, April 1990 参照)。図9に示すように左右各チャネルの信号に利得差、遅延差を与えた後、加算・減算して相関成分・相関差成分を生成する。これらは CELP 等のモノラル音声符号化を施した後、利得差、遅延差からなる相関付加成分とともにパケット伝送する。

【0094】復号側では、図10に示すように相関成分・相関差成分を各々復号した後、加算・減算を行い、付加情報として伝送された相関成分を用いて利得・遅延を復元して左右の信号を生成する。

【0095】システム面から検討する。

【0096】臨場感通信においては 7 KHz 以上の広帯域音声が見たい。そこで、図11に示すように帯域分

割処理は例えば4 KHz毎のサブ帯域に分割して処理する。各帯域の和成分は通常モノラル音声であるため固定レートで符号化し、差成分は複数発言時にのみ情報量が増大するため可変レートで符号化する。付加情報は和成分とともに伝送する。また、受話器をもたずに通話できる拡声通話は臨場感通信において重要な機能である。この場合、エコーキャンセラはスイッチ感のない自然な双方向通話が実現できる反面、処理量が多い。特に、ステレオ系にそのまま適用するとモノラルの4倍の処理量が必要になりコスト上不利である。そこで、図12に示すように推定伝達関数分解・合成型エコーキャンセラ

(S. MINAMI, "An Acoustic Echo Canceller For Pseudo-Stereophonic Voice" IEEE GLOBECOM '87 35.1 Nov. 1987 及び南 重信, "テレビ会議・マルチメディア端末音声系に適用する推定伝達関数分解合成型エコーキャンセラ", 電子情報通信学会 第6回 回路とシステムワークショップ pp. 333. 338 参照)を左右の相関成分に適用し、比較的電力の低い相関差成分に音声スイッチ方式を適用してモノラル並の処理量で実現できる。

【0097】通信互換性・セルロスを検討する。

【0098】広帯域音声端末と狭帯域音声端末との通信、ステレオ端末とモノラル端末との通信等、多様な端末間の相互通信を可能にする通信互換性の確保はマルチメディア通信における重要な課題の一つである。また、トラフィック時に発生するセルロス対策も考慮する必要がある。そこで、図13に示すように例えば通常のモノラル端末は伝送されたパケットのうち相関成分をそのまま復号すれば通常のモノラル音声を生成できる。また、相関成分と相関付加情報を優先パケットとして伝送し差成分を非優先とすると、差成分が輻輳等で消失しても、最低限、疑似ステレオ音声として生成できる。

【0099】映像符号化との連動を検討する。

【0100】本発明を適用したシステムでは、検出したチャンネル間相関より発言者位置が検出でき、発言者をより高精細に符号化したり発言者のウィンドウを抽出したりできる。

【0101】以上、チャンネル間相関を利用した広帯域ステレオ音声符号化方式、および本方式を用いた音声系のシステム構成について述べた

【0102】

【発明の効果】本発明によれば、音源が単独発言の時は伝送情報量の少ない疑似ステレオ音声符号化を、音源が複数発言の場合には従来通りの完全ステレオ音声符号化が適用でき、効率的なステレオ音声符号化が実現できる。また、復号側では送られてきた情報が疑似ステレオ

音声でも、完全ステレオ音声でも最適な復号ができる。

【0103】さらに、従来、処理量がモノラルの4倍必要であったステレオ音声用ECも疑似ステレオ用の推定伝達関数分解・合成形ECを用いることで、モノラルの2ないし4倍の処理量で効率的に実現できる。

【0104】このように本発明を用いれば、従来、困難であったステレオ音声を遠隔会議システムやマルチメディア端末にローコストで導入でき、臨場感や発言者の識別の向上に大きく貢献する。

10 【図面の簡単な説明】

【図1】本発明が適用されるマルチメディア会議システムの構成を示す図。

【図2】マルチメディア会議システムにおける書類の流れを示す図。

【図3】ステレオ-疑似ステレオ-モノラル音声間の通信互換性を確保する一実施例を示す図。

【図4】図3に示す第2のエコー抑圧手段の具体例を示す図。

【図5】図3に示す第2のエコー抑圧手段の他の具体例を示す図。

【図6】マイクアレイに遅延制御を適用した場合の構成例を示す図。

【図7】推定伝達関数分解フィルタの構成例を示す図。

【図8】安定性を保証した処理の構成例を示す図。

【図9】他の実施例に係る符合化装置の構成例。

【図10】他の実施例に係る復合化装置の構成例。

【図11】他の実施例に係る符合化方式を示す図。

【図12】他の実施例に係るECの構成例。

【図13】他の実施例に係るセルロス対策の説明図。

30 【図14】テレビ会議システムの概略図。

【図15】マルチメディア端末の概略図。

【図16】完全ステレオ音声系のシステムにECの技術を採用した例を示す図。

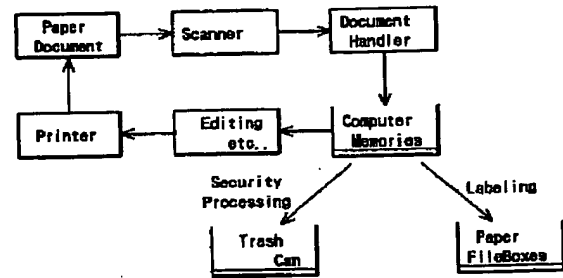
【図17】疑似ステレオ音声系にECの技術を採用した例を示す図。

【図18】分解合成型エコーキャンセラの概略を示す図。

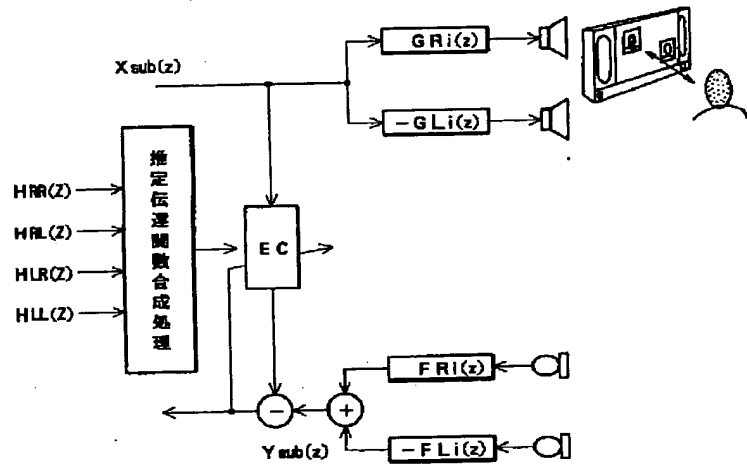
【符号の説明】

101-1、101-2…マイク、102-1、102-2…遅延・利得付与部、103、104、110-1…加算器、105、110-2…減算器、107…音響用エコーキャンセラ、108…遅延推定部、111-1、111-2…疑似ステレオ合成部、113…第2のエコー除去部、114…送信制御部、115…電力検出器。

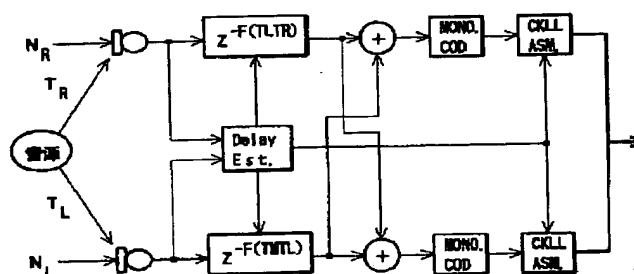
【図 2】



【図 15】

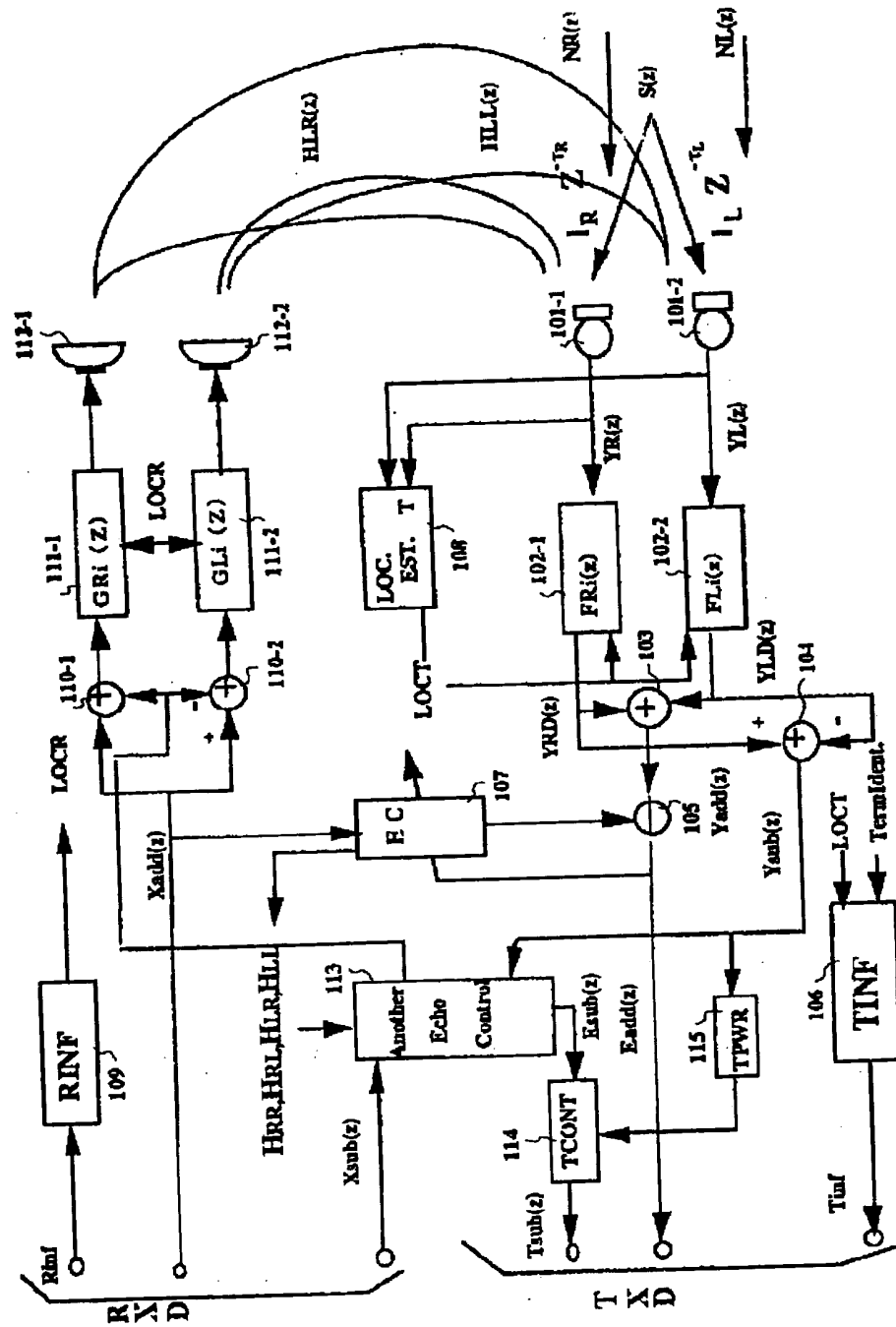


【图9】

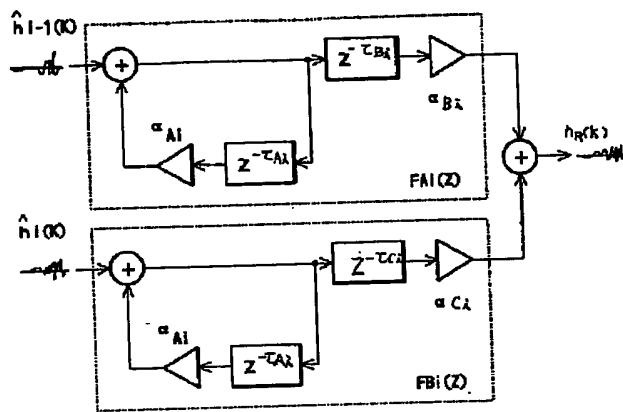


- 11 -

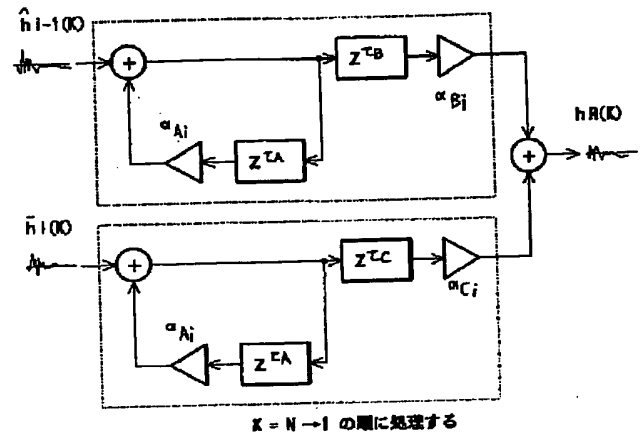
【図 3】



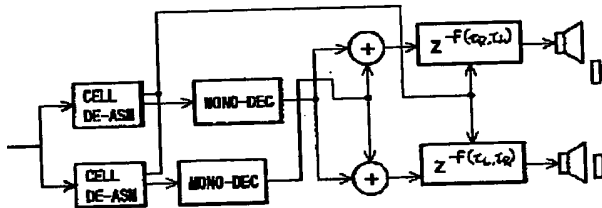
【図 7】



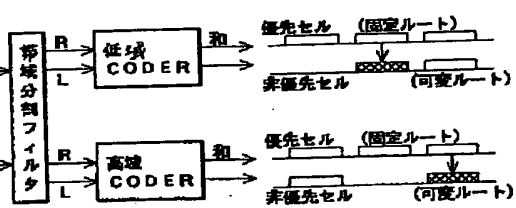
【図 8】



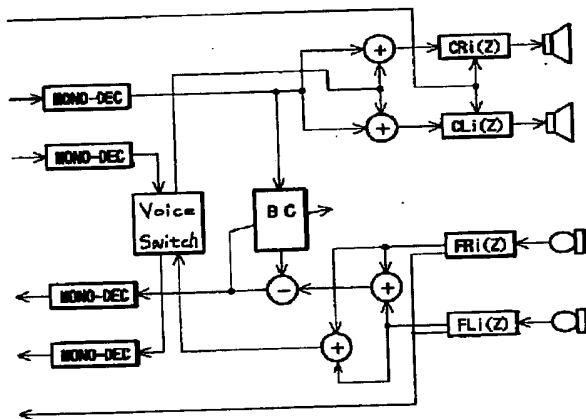
【図 10】



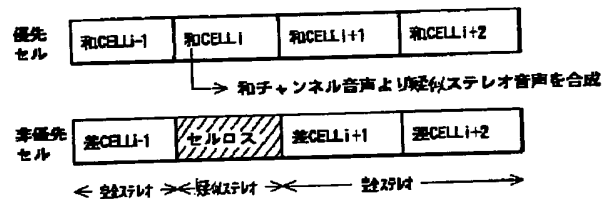
【図 11】



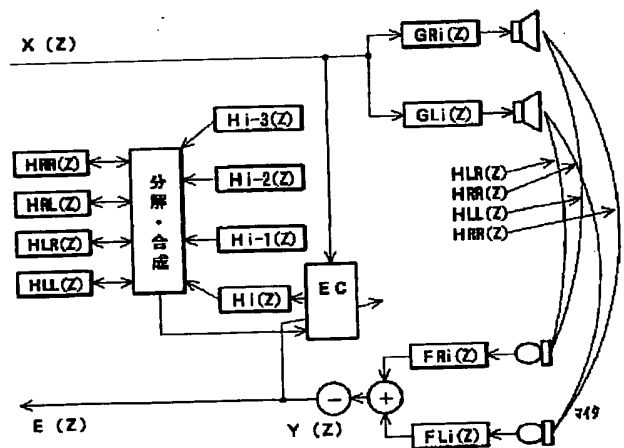
【図 12】



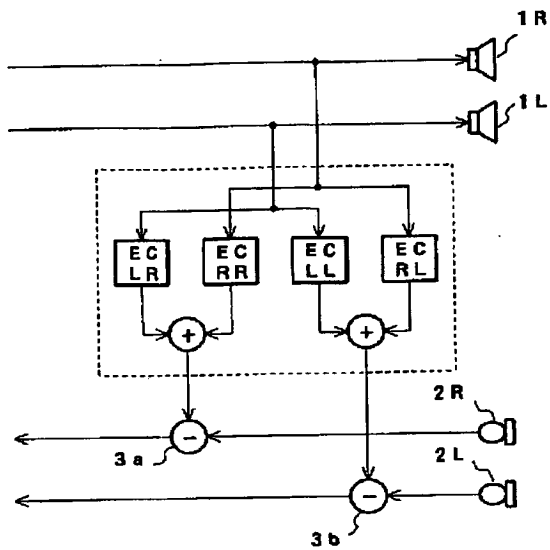
【図 13】



【図 18】



【図16】



【図17】

